Temporal and spectral cue use for initial plosive voicing perception by hearing-impaired children and normal-hearing children and adults

Lisa D. Holden-Pitt
Center for Auditory and Speech Sciences, Gallaudet University, USA

Valerie Hazan
Department of Phonetics and Linguistics, University College London, UK

Sally G. Revoie, Donna M. Edward and Janet Droge
Center for Auditory and Speech Sciences, Gallaudet University, USA

ABSTRACTS

The influence of voice-onset time (VOT) and vowel-onset characteristics on the perception of the voicing contrast for initial plosive consonants was examined for hearing-impaired children, and normal-hearing children and adults. Listeners identified spoken 'DAD'–'TAD' stimuli controlled for VOT and vowel onset characteristics. Only six of 16 hearing-impaired children appropriately identified the exemplar DAD and TAD stimuli used as endpoints of VOT continua. For this group of six hearing-impaired children, a longer VOT than for the normal-hearing listeners was required to elicit /t/ rather than /d/ percepts. The VOT region of perceptual cross-over in labelling widened progressively from normal-hearing adults to normal-hearing children to hearing-impaired children. Generally, longer VOTs were required to yield /t/ perception in the context of the DAD vowel than with the TAD vowel. These 'vowel stem' effects on VOT boundary were inconsistent for the hearing-impaired children, and weaker for the normal-hearing children than for the adults. These spoken stimuli produced results for VOT cue use that generally parallel those obtained in studies with synthetic stimuli.

L'on a examiné l'influence du délai de voisement (en anglais: le VOT) et des caractéristiques de l'attaque des voyelles sur la perception du contraste de voisement des occlusives initiales chez des enfants déficients auditifs, et des enfants et des adultes à l'audition normale. Les auditeurs ont dû identifier des stimulus 'DAD'–'TAD' prononcés, dont les caractéristiques de délais de voisement et d'attaque des voyelles étaient contrôlées. Seuls 6 des 16 enfants déficients auditifs ont identifié correctement les stimulus exemplaires DAD et TAD utilisés comme extrémités des continuums des délais de voisement. Pour ce groupe de 6 enfants, il a fallu un délai de voisement plus long que pour les auditeurs normaux pour obtenir des perceptions de /t/ plutôt que de /d/. La zone des délais de voisement résultant en un passage percepitif de l'un à l'autre s'est élargie progressivement des adultes jouissant d'une audition normale aux enfants jouissant d'une audition normale jusqu'aux enfants déficients auditifs. En général, il a fallu des délais de voisement plus longs pour obtenir une perception de /t/ dans le contexte de la voyelle de DAD que dans la voyelle de TAD. Ces effets de 'racine vocalique' sur la frontière de délai de voisement n'étaient pas systématiques chez les enfants déficients auditifs, et ils étaient plus faibles chez les enfants à l'audition normale que les adultes à l'audition normale. Les résultats de ces stimulus parlés pour l'utilisation du repérage du délai de voisement correspondent en général à ceux obtenus dans des études avec des stimulus synthétiques.

Key words: plosive voicing perception, hearing impaired, voice-onset time, acoustic cues.

INTRODUCTION

Study of the use of acoustic cues to the perception of consonant contrasts by hearing-impaired listeners provides insight into the nature of auditory deficits underlying reduced speech perception by this population. The use of auditory cues for the voicing contrast is particularly important, as this contrast is not cued visually. Phonemic feature distinctions may be elicited by two or more cues representing different acoustic domains. For example, voicing distinctions for initial plosive consonants can be cued by the duration of the interval between the onset of the plosive release burst and the onset of the glottal signal, i.e. voice-onset time (VOT) (e.g. Lisker, 1975b), and by the associated transition in the first formant (F1) at the vowel onset (e.g. Revoile, Pickett, Holden-Pitt, Talkin & Brandt, 1987).

VOT is typically two-to-three times longer for voiceless than for voiced plosives in American English (Stevens & Klatt, 1974). Spectral voicing cues may be evident in the vowel onset, where the F1 onset frequency for initial voiced plosives is generally lower than that for voiceless plosives. It is important to estimate the relative perceptual weighting of these different cues for hearing-impaired listeners, to ensure that the most auditorily salient cues to the contrast are well reproduced by the user's hearing aid.

The use of acoustic cues for initial plosive voicing distinctions by severely/profoundly hearing-impaired listeners has received a good deal of study. Greater importance of temporal than spectral cues for this phoneme distinction has been reported for severely hearing-impaired adults (Revoile et al., 1987). Among studies with hearing-impaired children, however, the relative perceptual importance of temporal versus spectral cue hierarchy is unclear.

In studies of children with different degrees of hearing impairment, those with moderate or mild hearing losses showed normal use of the cues for initial plosive voicing distinctions (Parady, Dorman, Whaley & Raphael, 1981; Johnson, Whaley & Dorman, 1984). Among the severely hearing-impaired children tested, listeners were about evenly divided between those showing normal versus subnormal performance in labelling plosive consonants that differed in voicing. Although most of the profoundly hearing-impaired children were unable to label the stimuli, several others with this degree of loss evidenced
voicing boundaries for some, if not all, of the stimuli tested. These results were obtained for synthetic CV stimuli in continua across which VOT increased systematically with concomitant cutback in F\textsubscript{1} at the vowel onset. Because both VOT and F\textsubscript{1} co-varied within the continua, the relative contribution of each cue to the voicing distinction was not determined.

In longitudinal investigations of the acquisition of the voicing contrast in initial plosive consonants, Hazan and Fourcin (1985) found dependence on the F\textsubscript{1} onset cue for some hearing-impaired children, whereas others relied on the VOT cue. They tested severely/profoundly hearing-impaired children between 7 and 13 years of age on a synthesised ‘goat–coat’ contrast that varied either for VOT alone or for the combination of VOT and vowel F\textsubscript{1} onset frequency. Some children who were able to label the combined-cue condition gave random responses when the F\textsubscript{1} onset cue was removed, suggesting a reliance by these children on the F\textsubscript{1} onset for cueing initial plosive voicing. Other children showed little difference in labelling between the combined-cue and ‘reduced-cue’ (i.e. neutralised F\textsubscript{1} onset frequency) conditions, an indication that these subjects were relying primarily on the VOT cue. Finally, a number of children were unable to label any stimuli, and thus could use neither the VOT nor the F\textsubscript{1} cue.

In a subsequent study of 16 severely/profoundly hearing-impaired children between 7 and 8 years old over a 4-year observation period (Abberton, Hazan & Fourcin, 1990), nine of the 16 children continued to display random labelling of a synthesised ‘goat–coat’ voicing continuum throughout the testing period. Thus these nine children, by 12 years of age, could use neither the VOT nor F\textsubscript{1} cues for distinguishing voicing.

The studies described, using synthetic stimuli, demonstrated that about half of the severely/profoundly hearing-impaired children tested were able to use either or both cues for initial plosive voicing distinctions. In contrast, the single study that used spoken stimuli found that a group of severely/profoundly hearing-impaired children was unable to distinguish voicing for initial plosives (Bennett & Ling, 1973). Here, stimuli were unaltered spoken CVs produced with VOTs differing in approximately 20-ms steps between 0 and 120 ms. Bilabial, alveolar and velar plosives were represented among the CV stimuli, with V = /a/, /u/ and /i/. For a group of 10 normal-hearing children, VOT boundaries of 20–40 ms were established. However, the 10 severely/profoundly hearing-impaired children obtained ‘no true [plosive] voiced–voiceless boundary . . . [although] there was a quasi-boundary at 60 ms beyond which voiceless plosives were recognized as such at a better than chance level’.

The discrepant results for perception of consonant voicing distinctions by the children tested with the spoken stimuli (Bennett & Ling, 1973) could have been related to uncontrolled variables among those stimuli, because each token of the ‘continuum’ represented a different spoken utterance, whereas, when synthetic speech stimuli are used, all acoustic patterns apart from the one under investigation are kept constant. On the other hand, the results obtained with the synthetic stimuli could be unique to those stimuli. The acoustic cue in synthetically produced stimuli may be more salient than those represented in spoken stimuli. Although this may occur as a result of the use of exaggerated spectral values relative to the equivalent cues in spoken stimuli (e.g. more extensive formant transitions) (Nittrouer & Studdert-Kennedy, 1987), the
perceptual salience of a cue in synthetic stimuli may also be pronounced as a result of the simplicity of the signal.

It is therefore important to test results obtained with synthetic stimuli through the use of (digitised) spoken stimuli for which acoustic characteristics have been carefully measured. Perceptual results yielded by synthetic stimuli might not be replicated when more natural, spoken utterances are used as stimuli (Shinn, Blumstein & Jongman, 1985; Nittrouer & Studdert-Kennedy, 1986). Further, when children are the subjects, the completeness of natural speech may be preferable to synthetic stimuli when studying acoustic-cue use for consonant perception (Nittrouer & Studdert-Kennedy, 1987).

The purpose of this study was to examine the ability of hearing-impaired children with varying degrees of hearing loss to use VOT and vowel-onset cues for plosive distinctions when spoken CVCs, carefully altered for these cues, were employed as test stimuli. Performance by the hearing-impaired children was compared with that of normal-hearing children and adults. Use of two normal hearing groups provided the opportunity to examine children versus adults for their use of VOT and vowel-onset cues in spoken stimuli, representing an initial plosive voicing contrast, a comparison made thus far primarily for synthetic stimuli. As moderately hearing-impaired children have not been tested for distinctions of plosive voicing via natural speech, our listener group included such children, as well as those with more severe losses.

**METHOD**

**Stimuli**

Two naturally spoken utterances—a [tæd] and a [dæd] (male talker)—were taken from a pool of stimuli used previously in a study of plosive voicing distinction by hearing-impaired adults (Revoile et al., 1987). These two utterances were selected to differ minimally for the acoustic attributes of vowel duration, vowel amplitude, and F0, but to be representative of the pool of 20 stimuli in terms of their extent of F1 transition and VOT. The original VOTs were 15 ms for the DAD utterance (range across 10 tokens: 15–18 ms) and 70 ms for the TAD utterance (range: 59–88 ms). The vowel F1 onset transition was intended to be the primary acoustic distinction between the two utterances (Table 1). The extent of F1 onset transition also was representative of that seen in the pool of 10 DAD and 10 TAD utterances, and similar to figures quoted in other studies that have used spoken stimuli (Howell, Rosen, Lang & Sackin, 1992). The analogue-to-digital conversion of these two recorded utterances was performed using a 10-kHz sampling rate, with low-pass filtering at 5 kHz and 12-bit amplitude representation.

So that the relative contribution of the temporal (VOT) and spectral (vowel onset) cues could be investigated, two continua of [tæd]–[dæd] test stimuli were developed. Within continua, the stimuli varied for VOT exclusively. The [t] burst (i.e. transient plus aspiration) extracted from [tæd] was used to prepare the initial plosive bursts for all the stimuli. VOTs for the eight constituent steps of each continuum were derived by deleting successive sections from the terminal end of the original [t] burst. The segments to be adjusted were identified visually on waveforms of the digitised utterances. VOT was measured from the onset of the burst release; the vowel start was signalled by the start of wave-
Table 1: Acoustic measurements of the [t] burst and of the vowels in the spoken [taed] and [daed]

<table>
<thead>
<tr>
<th>Original syllable</th>
<th>Duration (ms)</th>
<th>Amplitude* (dB)</th>
<th>Onset $F_0$ (Hz)</th>
<th>Mid-vowel $F_0$ (Hz)</th>
<th>Onset $F_1$ (Hz)</th>
<th>Mid-vowel $F_1$ (Hz)</th>
<th>$F_2$</th>
<th>$F_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[taed]</td>
<td>324</td>
<td>47</td>
<td>112</td>
<td>98</td>
<td>444</td>
<td>506</td>
<td>1813</td>
<td>2513</td>
</tr>
<tr>
<td>[daed]</td>
<td>325</td>
<td>48</td>
<td>111</td>
<td>100</td>
<td>391</td>
<td>525</td>
<td>1781</td>
<td>2494</td>
</tr>
</tbody>
</table>

*Relative to an arbitrary reference.

form periodicity at the fundamental, $F_0$. The test continua comprised eight stimuli, across which VOT varied from 17 to 60 ms in 6-ms steps nominally. All syllable modifications were made through digital signal processing.

Between continua, stimuli differed for their 'vowel stem' characteristics (i.e. [æd] portion of the syllable). In the ‘TAD Stem’ continuum, the [æd] portion from the digitised [taed] utterance served as the stimulus base. The spectral cues that it contained were therefore appropriate for a voiceless consonant. The eight durations of [t] burst were adjoined to the [taed] vowel stem, thus producing the continuum stimuli. For the other continuum, ‘DAD Stem’, the constituents were developed from the digitised spoken [daed] utterance, from which the initial [d] burst had been removed and replaced by the same eight prepared [t] burst durations used in the TAD Stem continuum. In this continuum, the spectral cues contained in the vowel stem therefore signalled a voiced prevocalic consonant. Spectrograms of the continuum–endpoint stimuli (i.e. those with the shortest and longest VOTs in each continuum) are shown on the left- and right-hand sides, respectively, of Figure 1.

Subjects
Ten normal hearing children aged between 7 and 9 years, 16 hearing-impaired children aged between 7 and 12 years and 10 normal-hearing adults* served as

*Data for the 10 normal hearing adults were obtained by combining the results for two sets of five normal-hearing adult listeners tested in separate experiments. These data were combined to increase the statistical power for the analysis of the listener group effects. From an earlier experiment on initial plosive consonant voicing (Revoile et al., 1987), DAD–TAD labelling data for the DAD Stem and TAD Stem continua were available for five normal-hearing adults. For each of the two studies, the same two spoken utterances were used to develop the stimuli. However, one set of five listeners received these stimuli via the adaptive paradigm described in the text, whereas VOT use for the other set of five listeners was assessed by the method of constant stimuli. For each of the 10 normal-hearing adult listeners, labelling function parameters per continuum were extracted via the method described in the text. An ANOVA comparing VOT boundary data for the two sets normal-hearing adults (mean = 26.1 ms, s.d. = 1.7, and mean = 24.3 ms, s.d. = 1.7, respectively) indicated no significant effect of listener set ($F[1, 8] = 0.37, p = 0.56$), and no significant interaction of listener set and continuum ($F[1, 8] = 1.6, p = 0.24$). Thus, data for these two sets of normal-hearing adult listeners were combined for further analyses.
Figure 1: Spectrograms of the endpoint stimuli for the TAD Stem and the DAD Stem continua. Continuum stimuli with the shortest VOT are represented on the left-hand side of the figure; those with the longest VOT are shown on the right-hand side.
paid listeners. The hearing-impaired children (all but one, pupils at the Kendall Demonstration Elementary School of Gallaudet University) had impairments of sensorineural origin with three-frequency (0.5, 1.0, 2.0 kHz) average tone thresholds (3FA) ranging from 55 to 97 dB HL (ANSI, 1969) in the better ear for hearing; the median loss was 80 dB HL. Eight of these children had severe losses, five had profound losses, and three moderate. Flat and sloping audiometric configurations were represented within each of these categories of hearing impairment. Thresholds for the normal-hearing listeners were 10 dB HL or better between 0.25 Hz and 4 kHz. The normal-hearing children were those of staff and faculty members at Gallaudet University.

Procedure
The stimuli were presented in single-interval identification trials with response alternatives of ‘TAD’ and ‘DAD’. Orthographic labels and colour drawings of ‘TAD’ (tadpole caricature) and ‘DAD’ (father figure) along with the Manual Alphabet signed letters ‘t’ and ‘d’, respectively, were displayed on the touch-sensitive screen of a terminal coupled to a computer.

The order of presentation for the stimuli was determined by an adaptive testing procedure (Hazan, 1986), which was a modification of the parameter estimation (PEST) method used by Taylor and Creelman (1967). The modified PEST procedure (Hazan, 1986) was developed for the specific purpose of testing a listener population of varying ability and limited concentration span, such as hearing-impaired children. The algorithm of this procedure reduces the number of stimulus presentations when subjects are unable to perceive the stimulus contrast being tested. Thus, the listeners are not frustrated by a lengthy number of trials for stimuli that are indistinguishable (for details of the adaptive test procedure, see Hazan, Fourcin & Abberton, 1991).

The normal-hearing and hearing-impaired children listened in five 30-minute sessions occurring within a 3-week period. The first session included tests of tone thresholds and repeated identification trials of the DAD–TAD endpoint stimuli. For some listeners whose labelling of the continuum endpoint stimuli remained poor during the first listening session, additional practice identifying these stimuli was provided during the second session. By the third session, administration of the full test continuum was begun. The DAD Stem and TAD Stem continua were each tested four times in quasi-random order over the remaining sessions.

Stimuli were presented to the normal-hearing listeners at 75 dB SPL, and to each hearing-impaired child at the child’s most comfortable listening level (MCL). The MCL used for each session was assessed via an adaptive tracking procedure in which an [æ], gated from the spoken [tæd] and presented twice per trial, was judged by the listener to be ‘too soft’, ‘perfect’ or ‘too loud’. The level of the signal changed in 2-dB steps across several reversals, which occurred for responses of ‘too loud’ or ‘too soft’. MCL was the mean level of the ‘perfect’ responses calculated over the final four reversals. MCLs resulting from two to three presentations of this procedure were used to determine the stimulus presentation level for the identification tests.

Each listener was tested in an IAC audiometric booth, with the stimuli presented monaurally to the more sensitive ear for sound, through a TDH-39
earphone (MX 41/AR cushion) in a headset. Test instructions were presented via total communication (voice and sign language) for the hearing-impaired children. Stimulus presentation and response tallies were under computer control (DEC PDP-11/23 and 11/34). Test sessions were carried out under identical conditions for children and adults, except that: (1) the adults used two to three one-hour listening sessions, and (2) an experimenter was present as a monitor in the sound booth during testing of the children. All participants were compensated for taking part.

RESULTS

Across the four test presentations per continuum (i.e. TAD Stem and DAD Stem), a mean performance level of over 70% correct labelling of the continuum extremes (i.e. endpoint stimuli) was accepted as reasonable demonstration of perceptual differentiation of these stimuli. Of the 16 hearing-impaired children, only six met this criterion. These six hearing-impaired children, henceforth referred to as ‘successful’, are represented in the statistical analyses later presented. Moderate (n = 2), severe (n = 3) and profound (n = 1) hearing losses were represented among these children. Two of them (both severely hearing-impaired) had flat audiometric configurations, whereas configurations for the other four were sloping.

Audiometric characteristics of the six successful, followed by the ten unsuccessful hearing-impaired listeners (Tables 2 and 3) provide some explanation of the performance differences between the two subgroups. The mean three-frequency tone threshold for the unsuccessful subgroup was about 10 dB poorer than that for the successful subgroup. Further, a larger difference in tone sensitivity to the lower frequencies was evident between the subgroups; the successful subgroup’s mean tone threshold at 0.5 kHz (56 dB) was better than that for the unsuccessful subgroup (77 dB). However, among the successful hearing-impaired children, there were listeners with 3FAS comparable to or greater than some for the unsuccessful children. Among the ten hearing-impaired children classified as ‘unsuccessful’, two responded distinctively to stimuli differing for vowel stem, with a majority of TAD responses to the TAD Stem stimuli regardless of their VOT, and largely DAD responses to the DAD Stem stimuli.

**DAD–TAD labelling**

For each listener per continuum, the number of DAD responses at each step was computed over the four administrations, to obtain a mean identification function per DAD Stem and TAD Stem continuum. A maximum likelihood estimation (MLE) procedure (Bock & Jones, 1968) was used to fit a cumulative normal function (probit analysis) to each subject’s set of data per continuum.*

Two parameters were extracted to characterise each identification function: (1) the voicing boundary, calculated as the 50% point of the fitted labelling curve, and (2) the gradient of the fitted curve (expressed as probit units divided

*Analyses of the labelling boundaries were performed using data in the form of VOT steps, as defined in the method section. However, to allow for greater ease in comparing the results with previous findings, the categorical boundary values were converted to milliseconds for display in the tables and figures.
by the number of stimuli in the continuum. The voicing boundary is an important measure for assessing effects of the various cue adjustments. For example, a large perceptual effect of spectral cues within the vowel stem would be reflected in a clear shift in voicing boundary between the TAD Stem and DAD Stem conditions. The labelling function gradient may be used as an indication of labelling consistency.

Table 2: Summary of ages and test ear audiometric characteristics for the 'successful' subgroup of hearing-impaired children

<table>
<thead>
<tr>
<th>Subject no.</th>
<th>Age (years)</th>
<th>Tone thresholds (dB HL)</th>
<th>3FA* (dB)</th>
<th>MCL (dB SPL)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.25 kHz</td>
<td>0.5 kHz</td>
<td>1 kHz</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>57</td>
<td>70</td>
<td>83</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>55</td>
<td>55</td>
<td>86</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>42</td>
<td>48</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>9.5</td>
<td>17</td>
<td>33</td>
<td>64</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>68</td>
<td>78</td>
<td>98</td>
</tr>
<tr>
<td>6</td>
<td>11.5</td>
<td>59</td>
<td>54</td>
<td>93</td>
</tr>
</tbody>
</table>

Mean: 9.5 50 56 79 86+ 90 74+ 113
s.d.: 1.1 17 15 7 15 18 13 12

*A plus sign (+) after the three-frequency average (3FA) indicates that for at least one octave frequency no detection threshold could be attained at the maximum limit of the audiometer. Thus, the maximum output level was used in the calculation of 3FA.

Table 3: Summary of ages and test ear audiometric characteristics for the 'unsuccessful' subgroup of hearing-impaired children

<table>
<thead>
<tr>
<th>Subject no.</th>
<th>Age (years)</th>
<th>Tone thresholds (dB HL)</th>
<th>3FA* (dB)</th>
<th>MCL (dB SPL)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.25 kHz</td>
<td>0.5 kHz</td>
<td>1 kHz</td>
</tr>
<tr>
<td>1</td>
<td>7.5</td>
<td>90</td>
<td>80</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>62</td>
<td>75</td>
<td>83</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>75</td>
<td>78</td>
<td>92</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>74</td>
<td>75</td>
<td>87</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>71</td>
<td>72</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>31</td>
<td>65</td>
<td>69</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>75</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>8</td>
<td>8.5</td>
<td>89</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>68</td>
<td>74</td>
<td>85</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>66</td>
<td>77</td>
<td>89</td>
</tr>
</tbody>
</table>

Mean: 9.5 70 77 85 91+ 93+ 85+ 117
s.d.: 1.2 16 8 9 15 24 9 11

*A plus sign (+) after the three-frequency average (3FA) indicates that for at least one octave frequency no detection threshold could be attained at the maximum limit of the audiometer. Thus, the maximum output level was used in the calculation of 3FA.
Table 4 shows listener group means and standard deviations of the VOT boundaries and labelling function gradients per stimulus continuum. Repeated measures analyses of variance and contrasts among listener groups were carried out separately for the boundary and gradient data. Bartlett-Box $M$ and Cochran's $C$ were used to test for homogeneity of variance.

Table 4: Means and standard deviations* for /dæd/~/tæd/ VOT boundaries (in ms) and identification function gradients per listener group

<table>
<thead>
<tr>
<th>Subject group</th>
<th>TAD Stem Boundary</th>
<th>Gradient</th>
<th>DAD Stem Boundary</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults ($n = 10$)</td>
<td>25.2 (1.8)</td>
<td>-2.12 (0.69)</td>
<td>32.8 (2.3)</td>
<td>-1.53 (0.55)</td>
</tr>
<tr>
<td>Hearing children ($n = 10$)</td>
<td>28.7 (2.2)</td>
<td>-0.92 (0.52)</td>
<td>32.2 (3.2)</td>
<td>-1.03 (0.50)</td>
</tr>
<tr>
<td>Hearing-impaired children ($n = 6$)</td>
<td>33.3 (4.6)</td>
<td>-0.49 (0.15)</td>
<td>35.6 (4.3)</td>
<td>-0.57 (0.24)</td>
</tr>
</tbody>
</table>

*Standard deviations are given in parentheses below each mean.

Voicing Boundaries
A difference among listener groups was seen for the effect of vowel stem upon the dependent variable of VOT boundary, as evidenced by a significant interaction of continuum and listener group ($F[2, 23] = 5.85; p = 0.009$). As a result, vowel stem effects were tested in separate analyses for the three listener groups. Results of these analyses revealed later VOT boundaries with the DAD Stem than with the TAD Stem continuum for normal-hearing children ($F[1, 9] = 20.0, p = 0.002$) and adults $F[1, 9] = 134.1, p = 0.0$). Although results for both normal-hearing groups showed a significant change in the DAD-TAD boundary dependent upon the continuum vowel stem, an analysis of data for only the two groups of normal-hearing listeners revealed an interaction effect for listener group by continuum ($F[1, 18] = 11.98, p = 0.003$). The group mean boundaries in Table 4 reveal that this statistical result reflects the greater impact of continuum upon perceptual boundaries for the normal-hearing adults than upon those of the normal-hearing children.

For the hearing-impaired children, no significant difference in VOT boundary was observed between the DAD Stem and TAD Stem continua ($F[1, 5] = 1.74, p = 0.24$). Thus, the effect of vowel stem on the DAD-TAD voicing boundary differed according to the group's hearing status, with the normal-hearing groups showing the greater effect of continuum on the TAD-DAD VOT boundary.

The hearing-impaired group's relatively large standard deviations for the voicing boundaries demonstrate their greater-than-normal intersubject variability. To examine the intragroup differences for the hearing-impaired children graphically, the estimated voicing boundaries per continuum are plotted for individual listeners in Figure 2, along with results for the normal-hearing listeners. The degree of consistency for continuum effect across listeners within a group is readily apparent. The slope of the lines connecting the continuum boundary pairs illustrates the impact of vowel stem on voicing boundary as measured by VOT. These lines are positively sloped for all of the normal-hearing
listeners, with the adults showing steeper slopes than nearly all of the children. More varied are the patterns of performance seen for the hearing-impaired children, for whom the lines connecting continuum boundary locations range from very positively sloping, to conspicuously negatively sloping.

![Graphs showing VOT locations of /dæd/ and /tæd/ boundaries for TAD Stem and DAD Stem continua for different listener groups.](image)

**Figure 2:** For each of the three listener groups, individual listeners' VOT locations of /dæd/-/tæd/ boundaries for TAD Stem and DAD Stem continua are plotted to illustrate the shift in boundary with change in vowel stem.

**Identification Gradients**

The comprehensive analysis of function gradient data for the TAD Stem and DAD Stem continua by the three listener groups was thwarted by the failure of the data to satisfy assumptions of homogeneity of variance among the listener groups (Box's $M = 21.0$, $F[6, 3590] = 3.015$, $p = 0.006$). Consequently, continuum effects were tested within each listener group. No significant differences in labelling function gradients were found between the TAD Stem and DAD Stem continua for any listener group ($F[1, 9] = 0.61$, $p = 0.454$ for normal hearing children; $F[1, 5] = 1.30$, $p = 0.306$ for hearing-impaired children; $F[1, 9] = 3.84$, $p = 0.082$ for normal hearing adults).

To illustrate listener group differences for the labelling function gradient, Figure 3 displays each group's mean labelling function for the TAD Stem continuum. The DAD Stem labelling functions are not shown, as the group-wise pattern of mean labelling functions was similar between the TAD Stem and the DAD Stem continua. For the TAD Stem and for the DAD stem continua, results from *a priori* contrasts revealed labelling function gradients to be more gradual.
for the hearing-impaired children than for the normal-hearing listeners ($t_{[23]} = -4.1$ and $-3.2$, respectively, $p < 0.01$), and steeper for the normal-hearing adults than for the normal-hearing children ($t_{[23]} = 4.9$ and $2.3$, respectively, $p < 0.05$).

![Diagram showing mean labelling functions for the three listener groups.]

**DISCUSSION**

This investigation assessed hearing-impaired children's use of VOT and/or vowel onset* as plosive voicing cues in spoken stimuli, in comparison to the use of these cues by normal-hearing children and adults. A secondary issue was the importance of vowel onset characteristics (in conjunction with VOT) for signalling the voicing contrast for normal-hearing children versus adults.

*Unsuccessful* Hearing-impaired Children

In comparison to studies of initial plosive voicing using synthetic stimuli (Parady et al., 1981; Johnson et al., 1984; Hazan & Fourcin, 1985), a higher proportion of our hearing-impaired children failed to categorise the continuum-endpoint stimuli developed from the spoken syllables. For some of these children, poor differentiation of the DAD–TAD stimuli probably resulted from insufficient burst audibility. However, in some cases this explanation of poor labelling performance was not tenable because their burst sensation levels (SLs) (estimated

*Although the contrast between continua is performed at the level of the complete vowel stem (i.e. vowel and final [d] burst), the research literature based on synthetic stimuli indicates that, for the voicing distinction, vowel onset is the critical characteristic. Also recall that our two base stimuli – [tæd] and [dæd] – were selected to differ minimally for attributes beyond the vowel F$_1$ onset transition. Thus, we occasionally refer to cues associated with the vowel stem as ‘vowel onset’ cues. However, without further alteration of the stimuli, we cannot be certain that vowel stem attributes external to F$_1$ onset are not significantly influencing the labelling performance patterns seen between continua.
from their selected listening levels relative to burst thresholds) well exceeded that typically needed to assure burst audibility. In other studies of plosive voicing distinction by hearing-impaired children (e.g. Parady et al., 1981; Johnson et al., 1984), degree of loss seemed to influence the use of VOT and vowel-onset cues. In the present study, however, degree of hearing impairment, per se, was not a primary determinant of the children's use of VOT for making the DAD–TAD distinction; the successful and the unsuccessful subgroups included listeners with moderate, severe and profound losses.

It should be noted that although the mean 3FA thresholds for the successful versus unsuccessful hearing-impaired listeners differed only somewhat, the successful subgroup displayed considerably better low-frequency hearing thresholds (see Table 2). Examination of the frequency composition of our spoken [t] burst revealed significant spectral energy at frequencies well below 1 kHz. Consequently, the relatively good low-frequency hearing of our successful hearing-impaired children probably resulted in greater audibility of the burst, facilitating its use for distinctions of plosive voicing. Hearing aid use, per se, did not appear to be the critical factor for successful VOT use: although all six successful children used their hearing aids each day all day, about half of the children in the unsuccessful subgroup were also daily hearing-aid users. Still, the fact that each child in the successful subgroup was a regular hearing-aid user should not be de-emphasised; evidence of adequate plosive burst sensation level appears to be a necessary, though not sufficient, condition for successful use of the VOT cue.

'Successful' Hearing-impaired Children

For the six hearing-impaired children able to distinguish [tæd] from [dæd], mean VOT boundaries occurred later in the continua than those for the normal hearing children and adults. Reduced burst audibility could explain the hearing-impaired group's relatively higher VOT boundary values. For our normal hearing listeners, DAD–TAD presentations at 75 dB SPL (for the vowel) probably yielded burst SLs of at least 35 dB, given that the average intensity of the [t] burst was 15 dB below that of the vowel. Although [t] burst thresholds were not measured directly, burst SL can be estimated for each hearing-impaired listener via spectral measurements (12.8-ms Hanning analysis window) of the burst and vowel segments, in conjunction with the listener's MCL for the vowel, and audiometric thresholds.* A listener's SL for a segment was estimated for the segment's spectral peak showing the maximum difference (in decibels) between presentation level (as determined relative to vowel F1 level) and threshold level. This estimation procedure resulted in a liberal measure of burst SL.

For the six successful hearing-impaired children, such estimates of burst SL ranged from 12 to 33 dB, with an average 24 dB – more than 10 dB below our

---

*To estimate the segment SLs, the following assumptions were made: (1) the strongest spectral prominence the vowel segment (i.e. F1) was presumably used by the subjects for determining comfort levels (MCL); and (2) a listener's thresholds at the octave frequencies nearest (in frequency) to the segment's respective spectral peaks reasonably approximate the listener's thresholds for peaks in that segment.
conservative burst SL estimate for the normal hearing listeners. The existence of a trading relation between VOT and amplitude of prevocalic plosive bursts (particularly aspiration) for cueing the voicing distinction has been suggested by Repp (1979), whose normal-hearing listeners required shorter VOT for distinguishing synthetic /ta/-/da/ as the aspiration-to-vowel intensity ratio was increased over a 24-dB range.

**Use of Vowel-onset Cues by Normal-hearing Adults and Children**

For the normal-hearing listeners, longer VOT boundary values were found for the DAD Stem stimuli (‘voiced’ vowel onset) relative to the TAD Stem stimuli (‘voiceless’ vowel onset). A significant shift in VOT boundary for the DAD Stem versus TAD Stem continuum is consistent with results obtained for normal-hearing listeners using synthetically produced stimuli, whereby stimuli with lower first formant (F₁) starting frequencies required longer VOTs to elicit voiceless plosive percepts than did stimuli with higher F₁ starting frequencies (Stevens & Klatt, 1974; Lisker, 1975a; Summerfield & Haggard, 1977).

Although some research suggests that normal-hearing children may be more sensitive than adults to spectral cues in the vowel onset (Morrongiello et al., 1984; Nittouer & Studdert-Kennedy, 1987; Nittouer, 1992), other findings indicate that children in the early stages of perceptual contrast development show less dependence than adults on vowel-onset cues (Simon & Fourcin, 1978). The ‘vowel stem’ effect seen for our normal-hearing adults was more dramatic than that for our normal-hearing children (see Figure 2). At least two interpretations of this finding might be proposed. First, the adults’ greater VOT boundary differential between the continua might indicate that they are affected more by change in vowel onset than the children. However, a closer look at the results reveals an alternative explanation. The normal-hearing children’s later (than the adults’) VOT boundary in the TAD Stem continuum appears to be responsible for the younger listeners’ more proximate VOT boundaries between continua. This finding seems to indicate the normal-hearing children’s greater (relative to the adults) perceptual weighting of the dynamic spectral cue, F₁ movement, albeit slight, in our TAD vowel onset. The latter rationale is consistent with that expressed in Morrongiello et al. (1984) and Nittouer and Studdert-Kennedy (1987) who, using synthetic speech stimuli, reported that children showed greater sensitivity than adults to spectral cues in the vowel transition. However, differences in specific stimuli characteristics make it difficult to directly relate effects obtained between our study and that of Morrongiello et al. (1984). The Morrongiello stimuli show a greater distinction than do ours for F₁ onset transition frequency extent for their ‘strong’ (381 Hz in ‘strong’ DAY) versus ‘weak’ (181 Hz in ‘weak’ DAY) transition cue conditions, relative to our 62 Hz (TAD Stem) versus 134 Hz (DAD Stem) F₁ transition extents (calculated from values in Table 1). A closer comparison can be made with a study by Howell et al. (1992) which used spoken BAT-PAT stimuli with cross-spliced initial consonant and vowel stem portions (F₁ transition extent in BAT: 79 Hz). Howell et al. also found that 2- to 4-year-old children were more sensitive to changes in vowel stem than adults. It is interesting to note that the spoken stimuli used in the Howell et al. and our study show much less extensive F₁ transitions than the synthetic stimuli used in the other studies.
Use of Vowel-onset Cues by Hearing-impaired Children

Relative to the normal-hearing listeners, the hearing-impaired children, as a group, appear to be less sensitive to voicing cues in the vowel onset. This was indicated by the latter group's failure to show a significant difference in category boundary for the DAD Stem versus TAD Stem continua. However, it should be pointed out that one of the six successful hearing-impaired children did show a striking shift in voicing boundary (> 10 ms) in response to the vowel stem change between continua (Figure 2). In addition, some hearing-impaired children may be influenced more by vowel-onset than VOT cues. Recall that for two of the 10 unsuccessful hearing-impaired children, labelling responses appeared to be governed primarily by the vowel onset cues, as they gave predominantly voiced responses to stimuli of the 'DAD stem' continuum, and voiceless responses to the 'TAD stem' stimuli. Therefore, there appears to be greater variance in terms of cue dominance than for normal-hearing listeners. Although a greater proportion of the hearing-impaired children appear to use VOT (a temporal feature) as their predominant cue to the voicing contrast, some of these children are using the spectral information in the vowel onset as a major cue to the contrast. A similar finding was reported by Hazan et al. (1991) in their longitudinal study.

Listener-group Differences: Hearing-impaired versus Normal Hearing

The variability in labelling the intermediate steps of the continuum, as evidenced by mean identification function gradients, was greater for the hearing-impaired group than for the normal-hearing groups. Listener group differences for VOT boundary location and variability have been reported by Johnson et al. (1984) and Parady et al. (1981). However, when compared with those results, the performance patterns observed in the present study have not been associated as closely with the listeners' severity of hearing impairment. Among our listeners, moderate, as well as more severe, hearing impairments yielded generally later VOT boundaries, and considerable variability in labelling.

Listener-group Differences: Normal-hearing Children versus Adults

The group of normal-hearing children, while showing /d/-/t/ VOT boundaries similar to those for the adults, were less consistent than the adults in their labelling of the stimuli. This is evident from the shallower gradients seen for the normal-hearing children's labelling functions. The normal-hearing children's reduced labelling consistency suggests that their ability to label stimuli differing in VOT and/or vowel onset had not yet stabilised. For normal-hearing children, studies of plosive voicing perception have tended to focus on the developmental acquisition of cues (Zlatin & Koenigskeche, 1975; Simon & Fourcin, 1978; Burnham, Earnshaw & Quinn, 1987). It is generally accepted that by age 7 years or so, the VOT boundary is rather similar to the adult boundary value, although response variability in the vicinity of the phonemic boundary is typically greater for the children. Simon and Fourcin (1978) outlined three stages in the development of voicing labelling behaviour among children aged from 2 to 14. Labelling consistency, represented here by the identification function gradient, increased with age up to about 12 years. Thus, the shallower identification functions obtained for our children, relative to our
adults, is a probable consequence of their lower level of perceptual development.

**Synthetic versus Altered Spoken Stimuli**

The altered spoken stimuli produced results comparable to those found in studies employing synthetic stimuli. The voicing boundaries averaged around 30 ms, similar to those for synthetically produced stimuli used by others (Lisker, 1975a; Parady et al., 1981; Johnson et al., 1984; Hazan & Fourcin, 1985). Also, the adult group's shift in the labelling boundary between the TAD Stem and the DAD Stem continua was about 8 ms. This amount of shift is proximate to the 10 ms reported by Lisker (1975a) for synthetic VOT continua differing for the onset frequency and/or transition presence of F₁.

For the hearing-impaired children successful in using VOT for voicing distinctions, these spoken stimuli yielded boundaries, labelling stability, and intersubject variability much like those reported in studies with synthetic speech stimuli. Note, however, that our spoken [t] burst contained substantial energy in the low frequencies, a characteristic not usually represented in synthetically produced plosive bursts. In the synthesis of plosive bursts, the cutback of F₁ typically has precluded the presence of low-frequency aperiodic energy in the burst. The hearing-impaired children's better hearing for lower than for higher frequencies generally results in greater dependence on spectral information contained in the low-frequency region. Thus, for these hearing-impaired listeners, the spectral omission in the F₁ region of synthetic plosive bursts might yield less audible synthetically produced (relative to spoken) bursts, potentially depressing use of the VOT cue, and resulting in an unrealistic representation of the hearing-impaired listeners' perceptual abilities. Examination of an intervocalic [t] burst produced by a different talker revealed even greater energy in the F₁ region than found in the burst from our syllable-initial [t]. However, this potential effect of this spectral omission in the synthetic bursts could be camouflaged by the listener's ability to rely on the cutbacks in F₁ onset that have generally paralleled VOT adjustments in reported studies where synthetically produced stimuli were used.

The likelihood of such greater use of the F₁ onset cue in synthetically produced plosives is increased, given the seemingly 'enhanced' F₁ transition frequency extents used in the synthetic relative to their spoken counterparts. The spectral pattern in these /t/ bursts is not representative of the typical 'diffuse rising' alveolar burst spectrum presented in the literature (e.g. Blumstein & Stevens, 1979).

**CONCLUSION**

Our results for normal-hearing listeners support previous findings with synthetic stimuli: children and adults generally required a longer VOT to recognise a stimulus as voiceless in the presence of the DAD vowel stem, that is, lower F₁ onset frequency. Relative to the children, the normal-hearing adults gave sharper categorisation of the stimuli and showed a greater change in labelling boundary in response to the change in vowel stem (i.e. TAD Stem versus DAD Stem).

A high proportion of hearing-impaired listeners was unable to categorise stimuli differing in VOT. Those who could: (1) required generally longer VOTs
than the normal-hearing listeners to elicit a voiceless percept, and (2) were not significantly affected by vowel stem manipulation. Also, the hearing-impaired listeners performed more heterogeneously than did the normal-hearing listeners.

Much of the research on acoustic cue use by hearing-impaired listeners has employed synthetically produced rather than spoken speech stimuli. The acoustical richness of spoken stimuli allows for greater practical generalisation of findings, though the ability to isolate specific parameters experimentally for determination of cue impact (as is possible in synthetic stimuli) is compromised. When using spoken utterances to test the significance of identified acoustic cues, it is important to minimise acoustic stimulus variability which might confound the results. Careful selection of stimuli, based on adequate measurement of the experimental variables as well as the non-manipulated acoustic parameters, is vital. Insufficient regard to this detail is a concern in the results of Bennett and Ling (1973). As each stimulus in their VOT ‘continuum’ represented a different spoken token, variability in non-experimental stimulus parameters might explain the inability of their hearing-impaired children to distinguish voicing via the VOT cue.

In some cases, gains achieved from the use of spoken stimuli may justify sacrificing total stimulus certainty. For instance, in our research, some hearing-impaired children seemed to depend on low-frequency energy in the spoken burst. This reliance might not have been apparent from studying their perception for plosive voicing via synthetic stimuli. The absence of this energy in the F1 region of the synthetic plosive bursts could result in an underestimation of hearing-impaired listeners’ full capability for VOT cue use. Perhaps a two-stage process of investigation is indicated: first test with synthetically produced utterances, which allow for greater stimulus control; then follow up with a reasonably close match of spoken stimuli (altered as necessary) to confirm generalisation of the earlier results. This approach would allow the shortcomings of one method to be compensated for by the other.

ACKNOWLEDGEMENTS
We are grateful for the cooperation of the administration, and the research, audiology and teaching staff members at the Kendall Demonstration Elementary School, without whom this study would not have been possible. We are also indebted to our former staff member Tim Bunnell, who developed the software for making the acoustical measurements of our stimuli. Our appreciation is extended to Joanna Bezek, James M. Pickett, Lisa Nelson, Beth McLeod, Tim Bunnell, Linda Kozma-Spytek and Ruth Boyd for their assistance in various stages of this manuscript’s preparation. This study was supported by a Gallaudet University Presidential Award, the Gallaudet Research Institute, and a grant from the NINCDS (NIH). Janet Droge is currently affiliated with Health and Education Resources, Inc. in Arlington, Virginia.

REFERENCES


